

ABSTRACT

FUNCTIONALS OF EXPERIMENTS.

by

Erik Torgersen

Institute of Mathematics, University of Oslo,
Norway

We shall consider representations of real valued functions of types of experiments which carry relations (structures) of experiments into analogous relations (structures) for numbers. It is shown that if the parameter set is finite then functionals which are affine for experiment mixtures and also continuous for the experiment topology may be expressed in terms of continuous non negatively homogeneous functions on the likelihood space. Assuming such a representation we characterize those functionals which are e.g. monotone for increasing information or multiplicative for experiment multiplication.

REPRESENTATIONS OF FUNCTIONALS OF EXPERIMENTS

by

Erik Torgersen

University of Oslo, Norway

Introduction

We shall in this paper consider functionals of experiments which reflect one or more of the well known structures for experiments. Before proceeding let us briefly summarize some of the concepts and results which we shall need to refer to. The fundamental reference on statistical experiments as they are discussed here is LeCam 1986. Torgersen 1976 is however adequate for a reading of this paper. It should be pointed out that an important technical feature of the theory of LeCam is that general problems concerning experiments having infinite parameter sets often may be reduced to the corresponding problem for experiments having finite parameter sets.

We shall here employ the traditional notion of an experiment E as a family $E = (P_\theta : \theta \in \Theta)$ of probability measures on a common measurable space, the sample space of the experiment. The set Θ is of course the parameter set of E .

If $E = (P_\theta : \theta \in \Theta)$ and $F = (Q_\theta : \theta \in \Theta)$ are experiments having the same parameter set Θ and if $\epsilon > 0$ then E is ϵ -definert w.r.t. F if and only if $\sup_{\theta} ||P_\theta T - Q_\theta|| < \epsilon$ for a transition T from E to F .

If the sample space of F is Euclidean and if E is dominated (or more generally coherent in the sense that bounded linear functionals on the L -space of E are representable as bounded measurable functions) then T may always be chosen as a Markov kernel. As shown by

LeCam this amounts to the condition that riskfunctions in F are ϵ -dominated by riskfunctions in E .

The smallest (it exists) constant ϵ such that E is ϵ -deficient w.r.t. F is called the deficiency of E w.r.t. F and is denoted as $\delta(E, F)$. Associated with the deficiency is the deficiency distance $\Delta(E, F) = \max(\delta(E, F), \delta(F, E))$. More generally we may consider the deficiency $\delta_k(E, F)$ for k -decision problems for $k = 1, 2, \dots, \infty$ with $\delta_\infty(E, F) = \delta(E, F)$.

If $\delta(E, F) = 0$ then we shall say that E is at least as informative as F and write this $E \succ F$. " \succ " is then a partial ordering. The corresponding equivalence relation say that E and F are equivalent (equally informative) if $\Delta(E, F) = 0$ i.e. if $E \succ F$ and $F \succ E$. For any experiment

$$M_i \prec E \prec M_a$$

where M_i is totally non informative and M_a is totally informative. Here M_i may be any experiment $(P_\theta: \theta \in \Theta)$ where P_θ does not depend on θ while M_a may be any experiment $(P_\theta: \theta \in \Theta)$ such that P_{θ_1} and P_{θ_2} are disjoint when $\theta_1 \neq \theta_2$. The quantities $\delta(M_i, E)$ and $\delta(E, M_a)$ as measures of information are studied in Torgersen (1981).

Two prominent methods of combining experiments are products and mixtures. Thus if $E = (P_\theta: \theta \in \Theta)$ and $F = (Q_\theta: \theta \in \Theta)$ then their product is the experiment $E \times F = (P_\theta \times Q_\theta: \theta \in \Theta)$. Choosing the experiments E and F with, respectively, probabilities $1-p$ and p we obtain the mixture $(1-p)E + pF = (1-p)P_\theta \oplus p Q_\theta: \theta \in \Theta$.

Clearly products and mixtures are commutative as well as associative up to the information preserving operations of grouping and permuting observations. In the same sense products is distributive w.r.t. mixtures.

A most convenient tool, introduced by Blackwell 1951, for studying experiments with finite parameter sets is their representations in terms of standard experiments and standard measures. The sample space of standard experiments with parameter set θ is the set K_θ of prior probability distributions on θ . Usually this set is identified with the fundamental probability simplex in R^θ . By this identification the one point distribution δ_θ in θ may be identified with the vector e^θ in R^θ whose θ' -th coordinate is 1 or 0 as $\theta' = \theta$ or $\theta' \neq \theta$.

A standard measure S (respectively standard probability measure) is a non negative measure S on K_θ such that $\int x(\theta) S(dx) \equiv 1$ (respectively $\int x(\theta) S(dx) = 1/m$ where m is the number of points in θ). If S is a standard measure and S_θ , for each θ , have density $x \rightarrow x(\theta)$ w.r.t. S then the experiment $(S_\theta: \theta \in \theta)$ derived that way is called a standard experiment. Thus a standard experiment is an experiment $(S_\theta: \theta \in \theta)$ on K_θ such that $x \rightarrow x(\theta)$ for each θ , is the density of S_θ w.r.t. $\sum_\theta S_\theta$.

Consider now any experiment $E = (P_\theta: \theta \in \theta)$ and put $f_\theta = dP_\theta / \sum_\theta P_\theta$ and $f = (f_\theta: \theta \in \theta)$. Thus the random variable f is the posterior distribution for the uniform priori distribution. The experiment $\hat{E} = Ef^{-1} = (P_\theta f^{-1}: \theta \in \theta)$ is then a standard experiment having $S = (\sum_\theta P_\theta) f^{-1}$ as its standard measure. \hat{E} is the standard experiment of E

while S is the standard measure of E . By sufficiency any experiment is equivalent to its standard experiment and standard experiments are equivalent if and only if they are equal. Thus experiments are equivalent if and only if they have the same standard experiments. In general experiments are equivalent if and only if their restrictions to common finite subparameter sets are equivalent. Thus we may, although equivalence classes of experiments are not well defined sets, consider the set of equivalence classes of experiments having the same general parameter set θ as well defined. The collection of experiments which are equivalent to a given experiment E is called the type of E .

We shall here limit our attention to properties and functions of experiments which respect equivalence ; i.e. are properties and functions of types. Dominatedness and separability for statistical distance respect equivalence. Minimal sufficiency and finiteness of sample space are examples of properties which do not respect equivalence. Dimension of sample space is not a functional of types while the power of the most powerful level α test for testing " θ_0 " against " θ_1 " respects equivalence. Other functionals which respects equivalence are provided by e.g. Fisher information, Kullback-Leibler information and the Hellinger transform. The latter functionals, in contrast to the test related functional, share the property of being affine under mixture.

There is a very general method for constructing such functionals (including those just mentioned) using non negatively homogeneous and measurable function h on the likelihoodspace $[0, \infty]^\theta$. This is done

for an experiment $E = (P_\theta : \theta \in \Theta)$ as follows: Choose any, say σ -finite, measure μ such that $P_\theta \ll \mu$ when $h(x)$ actually depend on the θ -th coordinate $x(\theta)$ of $x \in R^\Theta$. Then neither the existence nor the value of the integral $\int h(dP_\theta/d\mu : \theta \in \Theta)$ depend on how μ otherwise is chosen. This quantity is here, if it exists, denoted by $\int h(dE)$ or by $\int (dP_\theta : \theta \in \Theta)$ or by other suggestive notations.

Functionals which may be derived from non negatively homogeneous functions as just explained will here be called representable. Thus statistical distance, squared Hellinger distance and affinity between (for) P_{θ_0} and P_{θ_1} are representable. Indeed they may, respectively, be expressed as:

$\int |dP_{\theta_0} - dP_{\theta_1}|$, $\int (\sqrt{dP_{\theta_0}} - \sqrt{dP_{\theta_1}})^2$ and $\int \sqrt{dP_{\theta_0}} dP_{\theta_1}$. Generalizing the affinity we arrive at the Hellinger transform $H(\cdot|E)$ of the experiment $E = (P_\theta : \theta \in \Theta)$. This is the map which to each prior distribution

t on Θ with finite support assign the number $H(t|E) = \int_\Theta \Pi dP_\theta^{t_\theta}$.

Thus $H(t|\cdot)$ is representable for each prior t with finite support.

2. Classification of functionals

We shall here utilize well-known properties of standard experiments to describe certain important classes of functionals of experiments with a finite parameter set.

Consider a real valued functional Ω on the set of types of experiments with a, not necessarily finite, parameter set Θ . We shall say that:

- (i) Ω is non negative if $\Omega(E) \geq 0$ for all experiments E .
- (ii) Ω is monotonically increasing if $\Omega(E) \geq \Omega(F)$ when $E \succ F$.
- (iii) Ω is monotonically decreasing if $\Omega(E) \leq \Omega(F)$ when $E \succ F$.
- (iv) Ω is monotone if it is either monotonically increasing or monotonically decreasing.
- (v) Ω is convex if $\Omega((1-p)E + pF) \leq (1-p)\Omega(E) + p\Omega(F)$ for all $p \in [0,1]$ and all pairs (E,F) of experiments.
- (vi) Ω is concave if $-\Omega$ is convex.
- (vii) Ω is affine if it is both concave and convex.
- (viii) Ω is τ -continuous if it is continuous w.r.t. a given topology τ for the set of types.
- (ix) Ω is multiplicative if $\Omega(E \times F) = \Omega(E) \cdot \Omega(F)$ for all pairs (E,F) of experiments.
- (x) Ω is additive if $\Omega(E \times F) = \Omega(E) + \Omega(F)$ for all pairs (E,F) of experiments.

We may also consider maps from the set of types into other spaces than the real numbers. The terms introduced above for a real functional may then be used for general maps provided the corresponding terms have a "natural" interpretation in the range space of this map. Thus these terms are all well defined for maps from the set of types to any ordered topological vector space.

A map from the set of types of experiments with parameter set Θ into a set of functionals may also be called a transform.

Note that a map from the set of types of experiments with a finite parameter set θ into a topological space is, by compactness, convergence determining if and only if it is continuous and 1-1.

Let us consider a few functionals.

First of all the deficiency $\delta_k(E, F)$ is monotonically decreasing in \tilde{E} and monotonically increasing in F . The deficiency $\delta_k(E, F)$ as well as the deficiency distance $\Delta_k(E, F)$ are convex in each argument. This follows since $\delta_k(E, F)$ is, for fixed E (for fixed F), the supremum of affine functionals in $F(E)$. A study of $\delta(E, M_a)$ and $\delta(M_1, E)$ as measures of information may be found in Torgersen 1981. We have earlier encountered the functional $\Omega_\lambda \rightarrow 1 - \frac{1}{2} \delta(E, M_a | \lambda)$ before and have observed that this quantity is the maximal Bayes probability of guessing correctly the true distribution for the prior λ . If we put $\tilde{E} = (P_\theta : \theta \in \Theta)$ then $\Omega_\lambda(\tilde{E}) = \int \prod_{\theta} \lambda_\theta P_\theta = \int (\prod_{\theta} \lambda_\theta x_\theta) S(dx)$ where S is the standard measure of \tilde{E} . It was shown by Morse and Sacksteder, 1966, that the transform $\lambda \rightarrow \Omega_\lambda(\tilde{E})$ from K_Θ to $[0, \infty[$ is a characterization i.e. it determines the type of \tilde{E} . This may be seen by first extending the set of permissible distributions λ to the set of all non negative distributions on θ and then by showing that the right hand partial derivative of $\Omega_\lambda(\tilde{E})$ w.r.t. λ_{θ_0} is $\int_{A_{\theta_0}} x_{\theta_0} S(dx)$ where $A_{\theta_0} = \{x : \prod_{\theta} \lambda_\theta x_\theta = \lambda_{\theta_0} x_{\theta_0}\}$. Thus the quantity $P_{\theta_0} (dP_\theta / dP_{\theta_0} : \lambda_{\theta_0} / \lambda_\theta : \theta \in \Theta) = \int_{A_{\theta_0}} x_{\theta_0} S(dx)$ is determined when $\lambda_{\theta_0} > 0$ for all θ . This in turn implies that $\int_{P_{\theta_0}} (dP_\theta / dP_{\theta_0} : \theta \in \Theta)$ is determined for each $\theta_0 \in \Theta$. Let $(S_\theta : \theta \in \Theta)$ be the standard experiment of \tilde{E} so that $S = \sum_{\theta} S_\theta$ is the standard measure of \tilde{E} . If $\theta = \{1, \dots, m\}$, $B_\theta = \{x : x_1 = \dots = x_{\theta-1} = 0, x_\theta > 0\}$, and ϕ is bounded on compacts and positive homogeneous on $[0, \infty[$ then: $\int \phi(d\tilde{E}) = \int \phi(x) S(dx) = \sum_{\theta} \int_{B_\theta} \phi(x | x_\theta) S_\theta(dx)$ where the θ -th term for $\theta = \theta_0$ is determined by $\int_{P_{\theta_0}} (dP_\theta / dP_{\theta_0} : \theta \in \Theta)$. Thus, the transform $\lambda \rightarrow \Omega_\lambda(\tilde{E})$ determines the type of \tilde{E} .

We summarize some wellknown properties of this transform as well as of two other useful transforms in:

Theorem 1 (Hellinger, Morse-Sacksteder and norm transform).

Consider experiments with the same, not necessarily finite, parameter set θ . Let Λ and $L_f(\theta)$ denote, respectively, the set of probability distributions on θ with finite support and the set of measures on θ with finite support. Consider the following transforms:

The Hellinger transform which to each experiment $\mathcal{E} = (P_\theta : \theta \in \theta)$ associates the function $t \mapsto H(t|\mathcal{E}) = \int_{\theta} t dP_\theta$ on Λ .

The Morse-Sacksteder transform which to each experiment $\mathcal{E} = (P_\theta : \theta \in \theta)$ associates the function $\lambda \mapsto \|\sum_{\theta} \lambda_\theta P_\theta\|$ on Λ .

The norm-transform which to each experiment $\mathcal{E} = (P_\theta : \theta \in \theta)$ associates the function $a \mapsto \|\sum_{\theta} a_\theta P_\theta\|$ on $L_f(\theta)$.

Then each of these transforms determine the experiment up to equivalence.

Furthermore all transforms are affine, and induce the topology of Δ -convergence for restrictions to finite sub parameter sets. The Hellinger transform is multiplicative and monotonically decreasing while the other two transforms are monotonically increasing.

Referring to above mentioned works for details we may argue this as follows:

Proof: Firstly we may without loss of generality assume that θ is finite and then the proof of the first assertion for the norm transform is quite analogous to the proof just given of the same assertion for the Morse-Sacksteder transform. The fact that the Hellinger transform is determining may be derived from the uniqueness theorem for Laplace transforms. As the defining expressions

are all representable we conclude that all transformations are affine under mixtures. Continuity is a consequence of continuity of the representing functions. As the Δ -topology coincides with the topology of weak convergence for standard measures and since this topology is compact it follows that continuous determining transforms induces just that topology.

The statements on monotonicity follows from the fact that the representing functions are concave in the Hellinger case and convex for the other two transforms. Finally the multiplicity of the Hellinger transform follows by an easy application of Fubini's theorem. \square

We may also consider functionals which are additive for experiment multiplication. If Ω has this property then:

$$\begin{aligned} \Omega(E) + \Omega(M_a) &= \Omega(E \times M_a) = \Omega(M_a) \\ \text{and} \\ \Omega(E) + \Omega(M_i) &= \Omega(E \times M_i) = \Omega(E). \end{aligned}$$

It follows that the only real valued additive functional is the 0-functional. If, however, we permit infinite values then there are several interesting functionals which are both additive and affine as well as monotone. The most prominent examples are perhaps entropy (Kullback-Leibler information) and Fisher information. The first one is defined for dichotomies (P_1, P_2) and is the number $-E \log(dP/dP) = - \frac{d}{dt} [\int dP^{1-t} dP^t]$ where the derivative is the right side derivative at $t=0$. Fisher information is defined for experiments whose parameter set θ is a sub set of R^k . If $\theta \subseteq R^k$ and if $\theta \in \theta^0$ then, under regularity conditions, $\delta(1 - \int \sqrt{dP_\theta} dP_{\theta+h}) = h' I_\theta h + o(\|h\|^2)$ where I_θ is the Fisher information matrix at θ .

We shall not discuss these information concepts further here, but just note that their basic properties of additivity, affinity and monotonicity follow from, respectively, the multiplicativity, affinity and monotonicity of the Hellinger transform.

Another interesting and closely related area is the theory of majorization and Schur convexity. Schur convex functions (Schur increasing functions might have been a better term) are functions on R^n which are monotonically increasing for the ordering called majorization. This ordering is precisely the ordering of being "at least as informative for" a particular kind of dichotomies; or rather pseudo dichotomies since we should not require that our measures are probability measures. Thus Schur convex functions may, see Torgersen 1985, be considered as monotonically increasing functionals in essentially the same sense as considered here.

Before discussing functionals in general let us make a short excursion to a related problem. Let us consider a complete separable metric space (X, d) . Let \mathcal{P} denote the set of probability measures on the class of Borel subsets of X . Topologize \mathcal{P} by its weak topology i.e. the smallest topology which makes $P(f)$ continuous in P when $f \in C(X)$. If $f \in C(X)$ then the functional $P \mapsto P(f)$ is affine as well as continuous. The interesting fact is now, see Huber 1980, that there are no other functionals on \mathcal{P} which are both continuous and affine and, in fact, if Q is such a functional then $Q(P) \equiv \int P(f)$ where $f(x) \equiv Q(\delta_x)$. The function f is continuous as well as bounded. This may, the arguments are taken from Huber's book, be seen as follows:

Firstly if x_n is any sequence in X then $(1 - \frac{1}{n})\delta_x + \frac{1}{n}\delta_{x_n} \rightarrow \delta_x$. It follows that $\frac{1}{n}Q(\delta_{x_n}) \rightarrow 0$. Hence f is bounded and it is clearly continuous. Furthermore $Q(P) = \int P(f)$ when P has finite support. If $P \in \mathcal{P}$ then there are distributions P_n , $n = 1, 2, \dots$ with finite supports such that $P_n \rightarrow P$. Hence $Q(P) = \lim_n Q(P_n) = \lim_n \int P_n(f) = \int P(f)$.

We might try to mimic this procedure for functionals of experiments. There is, however, the difficulty that one point distributions are not standard probability measures. Fortunately there are particular experiments which we may utilize instead. If Θ is finite then these may be described as follows:

Let us take the set θ together with a point ξ (e.g. the set θ itself) which does not belong to θ . If $\xi \in [0,1]^\theta$ then we define the experiment $G_\xi = (P_\theta: \theta \in \theta)$ by putting $P_\theta(\theta) = 1 - \xi_\theta = 1 - P_\theta(\xi)$. The standard measure of G_ξ assigns then mass $1 - \xi_\theta$ to each vertex e^θ of K_θ and the remaining mass $\sum_\theta \xi_\theta$ is assigned to $\xi / \sum_\theta \xi_\theta$. Note that $G_\xi \sim M_a$ if and only if there is at most one θ such that $\xi_\theta > 0$ while $G_\xi \sim M_1$ if and only if $\xi = (1, 1, \dots, 1)$. If h is a positively homogeneous function on $[0, \infty]^\theta$ then $\int h(d_\xi) = \sum_\theta (1 - \xi_\theta) h(e^\theta) + h(\xi)$.

If $\xi, \eta \in [0,1]^\theta$ then $\xi \eta \in [0,1]^\theta$ and $G_{\xi \eta} \sim G_\xi \times G_\eta$. It follows that $G_\xi > G_\eta$ when $\xi < \eta$. If, on the other hand, $G_\xi \geq G_\eta$ then, by the sublinear function criterion, $(\xi_{\theta_0} - \beta \xi_{\theta_1})^+ - \xi_{\theta_0} > (\eta_{\theta_0} - \beta \eta_{\theta_1})^+ - \eta_{\theta_0}$ when $\theta_0, \theta_1 \in \theta$ and $\beta > 0$. $\beta \rightarrow \infty$ yields $\xi_{\theta_0} < \eta_{\theta_0}$ provided $\xi_{\theta_1} > 0$. Thus $G_\xi \geq G_\eta$ & $G_\xi \not\sim M_a \Rightarrow \xi < \eta$. Hence $G_\xi > G_\eta \Leftrightarrow G_\xi \geq G_\eta \Leftrightarrow \#\{\theta: \xi_\theta > 0\} = 1$ or $\xi < \eta$.

It may be shown that the class of experiments G_ξ is contained in the slightly larger class of experiments E such that $F > E$ if and only if $F \sim E \times H$ for some experiment H . Consider now any experiment E having standard probability measure \bar{S} . Let Q_ξ denote the standard measure of G_ξ . Then $T = \int Q_\xi \bar{S}(d\xi)$ is the standard measure of some experiment which we may suggestively write as $\int G_\xi \bar{S}(d\xi)$. If h is non negatively homogeneous measurable and bounded on compacts then $\int h dT = \int [\sum_\theta h(e^\theta)(1 - \xi_\theta) + h(\xi)] \bar{S}(d\xi) = \sum_\theta h(e^\theta)(1 - \frac{1}{m}) + \int h(\xi) \bar{S}(d\xi) = (1 - \frac{1}{m}) \int h(dM_a) + \frac{1}{m} \int h(dE)$ where $m = \#\theta$.

Thus

$$\int G_\xi \bar{S}(d\xi) \sim (1 - \frac{1}{m}) M_a + \frac{1}{m} E.$$

Let now Ω be any affine and continuous functional on the set of experiments with parameter set θ . Define the experiments G_ξ as above and let E be an experiment with standard measure S and standard probability measure $\bar{S} = S/m$ where $m = \#\theta$.

The decomposition given before theorem 2 yields:

$$\Omega(\int G_{\xi} \bar{S}(d\xi)) = (1 - \frac{1}{m})\Omega(M_a) + \frac{1}{m}\Omega(E).$$

Approximating E with experiments with finite sample spaces we find, by continuity, that the left hand side may be written $\int \Omega(G_{\xi}) \bar{S}(d\xi)$. Thus

$$\Omega(E) = \int [\Omega(G_{\xi}) - (1 - \frac{1}{m})\Omega(M_a)] S(d\xi).$$

This proves:

Theorem 2. (Representations of continuous affine functionals).

Assume that θ is finite. Then a functional Ω on the set of types of experiments is real valued continuous and affine if and only if Ω is representable as

$$\Omega(E) = \int h(dE)$$

for some continuous and positively homogeneous function h on $[0, \infty]^\theta$.

As mentioned in the introduction any positively homogeneous measurable function h on $[0, \infty]^\theta$ which is bounded from below by a linear function defines an affine functional Ω by:

$$\Omega(E) = \int h(dE).$$

If $h(x) \equiv \sum_{\theta} c_{\theta} x_{\theta}$ where $\sum_{\theta} c_{\theta} = 0$ (i.e. h is a contrast) then $\int h(d) = 0$. Thus the representing function h is not unique.

There is, however, no more arbitrariness than that since h is unique up to an additive contrast. To see this consider a function h so that $\int h(dE) = 0$ for all experiments E . Then

$\int_{\theta} h(e^{\theta})(1-\xi_{\theta}) + h(\xi) \equiv \int_{\xi} h(\xi) \equiv 0$. Thus $h(\xi) = \int_{\theta} h(e^{\theta})(\xi_{\theta}-1)$ when $\xi \in [0,1]^{\theta}$. $\xi = 0$ yields $0 = h(0) = - \int_{\theta} h(e^{\theta})$ so that h is a contrast.

Let us next consider those functions h which define non negative functionals.

Assume then that h is a real valued measurable positively homogeneous function (thus $h(tx) = th(x)$ when $t > 0$) on $[0, \infty]^{\theta}$ such that $\int h(dE) > 0$ for all experiments E . Then $h(\xi) + \int_{\theta} (1-\xi_{\theta})h(e^{\theta}) = \int h(d\xi) > 0$ when $\xi \in [0,1]^{\theta}$. It follows in particular that h is bounded from below on K_{θ} . Let \underline{h} denote the largest continuous convex function on K_{θ} which is majorized by h . Consider in particular the value of \underline{h} at $e/m = (\frac{1}{m}, \dots, \frac{1}{m})$. There are then, since $\{(x,y): y \geq \underline{h}(x)\}$ is the convex hull of $\{(x,y): y \geq h(x)\}$, vectors x^1, \dots, x^m in K_{θ} and weights $\lambda_1, \dots, \lambda_m$ such that $e/m = \sum_{i=1}^m \lambda_i x^i$ and $\underline{h}(e/m) = \sum_{i=1}^m \lambda_i h(x^i)$.

Let S be the standard probability measure which assigns mass λ_i to x^i , $i = 1, \dots, m$. If S is the standard probability measure of the experiment E then, since $E \succ M_i$, $\int \underline{h}(x) S(dx) \geq \underline{h}(e/m) = \int h(x) S_0(dx) > 0$. Thus \underline{h} also defines a non negative affine functional and $h \geq \underline{h}$. Extend \underline{h} to a sub linear functional on R^{θ} and let $c \in R^{\theta}$ define a supporting hyperplane of $\{(x,y): y \geq \underline{h}(x)\}$ at $(e/m, \underline{h}(e/m))$ i.e:

$$\underline{h}(x) \geq (c, x) \quad \text{for all } x \geq 0$$

while $0 \leq \underline{h}(e) = (c, e)$. We may now decrease c , this does not affect the first inequality, and thereby obtain a vector $a \in R^{\theta}$ such that:

$$\underline{h}(x) \geq (a, x) \quad \text{for all } x \geq 0$$

while $0 = (a, e)$. Altogether we have found a contrast, $x \mapsto (a, x)$ on R^{θ} such that $h(x) \geq (a, x)$ for all $x \geq 0$.

This proves the non trivial part of:

Theorem 3. (Non negative representable functionals).

If h is a measurable non negatively homogeneous function on $[0, \infty[^\theta$ which is bounded on bounded sets then the affine functional $E \mapsto \int h(dE)$ is non negative if and only if h is minorized by a contrast on R^θ .

This yields the following characterization of monotone representable functionals.

Corollary 4. (Monotone representable functionals).

If h is a measurable non negatively homogeneous function on $[0, \infty[^\theta$ which is bounded on bounded sets then the affine functional $E \mapsto \int h(dE)$ is monotonically increasing (decreasing) if and only if h is sub linear (super linear) on R^θ .

Remark. h need not be continuous. If we consider the restriction of h to K_θ then the theorem may be phrased: " h defines a monotonically increasing (decreasing) functional if and only if $h|_{K_\theta}$ is convex (concave).

Proof: The "if part" of the corollary was established by Blackwell 1951. Note next that if h is minorized by a contrast on R^θ and if $h(e^\theta) \equiv 0$ then this contrast is the zero contrast and thus $h > 0$.

Assume so that h defines a monotonically increasing functional. Let $v, w \in K_\theta$ and let $t \in [0, 1[$. Put $u = (1-t)v + tw$. Define a dilation D from K_θ to K_θ by:

$$\begin{aligned} D(x|x) &= 1 \quad \text{if } x \neq u \\ D(v|u) &= 1-t \\ D(w|u) &= t. \end{aligned}$$

Put $\tilde{h}(x) = \int h(y) D(dy|x) - h(x)$ when $x \in K_\theta$. Then $\tilde{h}(e^\theta) = h(e^\theta) - h(e^\theta) = 0$. If S is a standard probability measure we obtain:

$$\int \tilde{h}(x) S(dx) = \int h(y) (DS)(dy) - \int h(y) S(dy) > 0$$

since the experiment defined by DS is at least as informative as the experiment defined by S . Hence, by the previous result, $\tilde{h} > 0$. In particular $0 < \tilde{h}(u) = (1-t)h(v) + th(w) - h((1-t)v + tw)$. Hence h is convex. \square

Let us so consider representable and multiplicative functionals:

Thus $\Omega(E) = \int h(dE)$ where h is a positively homogeneous and measurable function on $[0, \infty[^\theta$. The assumption of multiplicativity implies, since $E \times M_1 \sim E$ and $E \times M_a \sim M_a$, that $\Omega(E)\Omega(M_1) = \Omega(E)$ and $\Omega(E)\Omega(M_a) = \Omega(M_a)$ for any experiment E .

Define Ω_0 and Ω_1 by: $\Omega_0(E) \equiv \int_E 0$ and $\Omega_1(E) \equiv \int_E 1$. Then Ω_0 and Ω_1 are both multiplicative. Ω_0 is representable by h if and only if h is a contrast while Ω_1 is representable by h if and only if $h(x) \equiv \int_\theta x_\theta / m + c(x)$ where $c(x)$ is a contrast in x .

Assume next that Ω is multiplicative and that $\Omega \neq \Omega_0$ and $\Omega \neq \Omega_1$. Then there are experiments E and F such that $\Omega(E) \neq 0$ while $\Omega(F) \neq 1$. The identities above imply then that $\Omega(M_1) = 1$ and $\Omega(M_a) = 0$. If Ω is represented by h then $\Omega(\mathcal{G}_\xi) = \int h(d\theta_\xi) = h(\xi) + \int_\theta (1-\xi_\theta) h(e^\theta)$. Adding a suitable contrast we

may arrange things so that $h(e^\theta) = \kappa$ does not depend on θ . Then $\Omega(G_\xi) = h(\xi) + \sum_{\theta} (1 - \xi_\theta) \kappa$. Putting $\xi = 0$ we find $0 = \Omega(M_a) = \Omega(G_0) = h(0) + m\kappa = m\kappa$ so that $\kappa = 0$. Hence $\Omega(G_\xi) = h(\xi)$ when $\xi \in [0, 1]^\theta$.

Let $x, y \in [0, \infty]^\theta$. Then there is a constant $t > 0$ so that $\xi = tx$ and $\eta = ty$ both belong to $[0, 1]^\theta$. Then $h(xy) = h(\xi\eta/t^2) = h(\xi\eta)/t^2 = \Omega(G_{\xi\eta})/t^2 = \Omega(G_\xi \times G_\eta)/t^2 = [\Omega(G_\xi)/t] \cdot [\Omega(G_\eta)/t] = [h(\xi)/t] \cdot [h(\eta)/t] = h(\xi|t) \cdot h(\eta|t) = h(x) \cdot h(y)$. Define, for each θ , the function h_θ on $[0, \infty[$ by:

$$h_\theta(z) = h(1, \dots, \overset{(\theta)}{z}, \dots, 1).$$

Then, since $x = \prod_{\theta} (1, \dots, \overset{(\theta)}{x_\theta}, \dots, 1)$, we find that $h(x) = \prod_{\theta} h_\theta(x_\theta)$.

Furthermore $h_\theta(z_1 z_2) = h(1, \dots, z_1 z_2, \dots, 1) = h(1, \dots, z_1, \dots, 1) h(1, \dots, z_2, \dots, 1) = h_\theta(z_1) \cdot h_\theta(z_2)$. In particular $h_\theta(z) = (h_\theta(\sqrt{z}))^2 > 0$ for all θ and $z > 0$. Thus $h > 0$. If $z_0 > 0$ and if $h_\theta(z_0) = 0$ then $h_\theta(z) = h_\theta((z/z_0)z_0) = h_\theta(z/z_0)h_\theta(z_0) = 0$ for all $z > 0$. Hence $h = 0$ contradicting the assumption that $\Omega \neq 0$. It follows that $h_\theta(z) = (h_\theta(\sqrt{z}))^2 > 0$ when $z > 0$. On the other hand $h_\theta(0) = h_\theta(0 \cdot 0) = h_\theta(0)^2$ so that $h_\theta(0) = 1$ or $h_\theta(0) = 0$. If $h_\theta(0) = 1$ then $h_\theta(z) = h_\theta(z) \cdot h_\theta(0) = h_\theta(z \cdot 0) = h_\theta(0) = 1$ for all $z > 0$. Put $\phi_\theta(x) = \log h_\theta(e^x)$ when $x \in \mathbb{R}$. Then

$\phi_\theta(x+y) = \phi_\theta(x) + \phi_\theta(y)$ when $x, y \in \mathbb{R}$. This, together with the measurability of ϕ_θ , implies that $\phi_\theta(x) \equiv t_\theta x$ for some real constant x . (Put $H = \{x: \phi_\theta(x) = x\phi_\theta(1)\}$. Then H is a measurable sub group of $(\mathbb{R}, +)$ which contains the rational numbers: ϕ_θ is continuous since it is measurable and midconvex, see e.g.

Roberts and Varberg 1973. Thus H is closed so that $H = \mathbb{R}$.) It follows that $h_\theta(z) = e^{\phi_\theta(\log z)} = z^{t_\theta}$ when $z > 0$. If $t_\theta < 0$ then $h(1, \dots, \overset{(\theta)}{z}, \dots, 1) = h_\theta(z) \rightarrow \infty$ as $z \rightarrow 0$. This contradicts the integrability of h (finiteness of Ω) w.r.t. any standard

measure. Thus $t_\theta > 0$. If $a > 0$ then $h(a, a, \dots, a) = \prod_{\theta} h_\theta(a) = a^{\sum_{\theta} t_\theta}$ and $h(a, a, \dots, a) = ah(1, 1, \dots, 1) = 1$. Hence $\sum_{\theta} t_\theta = 1$. What happens to h_θ at $z = 0$? If $t_\theta > 0$ then we can't have

$h_\theta(0) = 1$, since this implies that $h_\theta(z) \equiv 1$. Thus $h_\theta(0) = 0$ in this case so that $h_\theta(z) = z^{t_\theta}$ for all $z > 0$ when $t_\theta > 0$. If $t_\theta = 0$ then continuity dictates that we should have $h_\theta(0) = 1$. There is, however, no way to show that $h_\theta(z)$ is continuous at $z = 0$ when $t_\theta = 0$. In fact we are free (just check it) to interpret 0^0 as 1, which is customary, but also as 0. The first choice makes h continuous. This choice does not, however make $\int \prod_{\theta} z_\theta^{t_\theta} S(dz)$ continuous as a function of t throughout K_θ when the experiment E with standard measure S is non homogeneous. If we, on the other hand, interpret all expressions 0^0 as 0 then h is not continuous when $t = (t_\theta : \theta \in \Theta) \in \text{bd} K_\theta$. Now, however, $\int \prod_{\theta} z_\theta^{t_\theta} S(dz)$ becomes continuous in t for each standard measure S .

Altogether this proves:

Theorem 5. (Multiplicative representable functionals).

Let h be a non-negatively homogeneous real valued measurable function on $[0, \infty[^\Theta$ which is bounded on bounded sets and put $\Omega(E) = \int h(dE)$. Then Ω is multiplicative and constant if and only if either $\Omega = 0$ or $\Omega = 1$. The first condition is satisfied if and only if h is a contrast while the second condition is satisfied if and only if h differs from $x \mapsto (\prod_{\theta} x_\theta) / \prod_{\theta} t_\theta$ by a contrast.

The functional Ω is multiplicative and non constant if and only if there is a prior distribution t on Θ such that h differs by a contrast from a function $z \mapsto \prod_{\theta} z_\theta^{t_\theta}$ where the factor $z_\theta^{t_\theta}$ should, for each θ , be interpreted either as 0 or as 1 when $z_\theta = t_\theta = 0$.

Any such function h is concave (i.e. superadditive since h is homogeneous) on $[0, \infty[^\Theta$ and thus defines a monotonically decreasing functional.

Remark. The assumption which assured that $t_\theta > 0$ for all θ was the assumption that θ was real valued. If we permit the values $+\infty$ for Ω then $E \rightarrow \int h(dE)$ is multiplicative (and affine) for any function h on R^θ of the form $h(x) \equiv \prod_{\theta} x_\theta^{t_\theta}$ where the real constants $t_\theta: \theta \in \theta$ are only subject to the condition $\sum_{\theta} t_\theta = 1$.

If we agree to put $0^0 = 1$ for all factors 0^0 appearing in $h(x) \equiv \prod_{\theta} x_\theta^{t_\theta}$ then the map $t \rightarrow \int h(dE) = \int \Pi(dP_\theta)^{t_\theta}$ from K_θ to $[0,1]$ is the Hellinger transform $H(\cdot|E)$ of E as defined in chapter M. As $H(\cdot|E)$ is, for fixed E , an integral of log convex functions it is log convex. If we, on the other hand, interpret each factor 0^0 as 0 then we get another convex function $\tilde{H}(\cdot|E)$ on K_θ .

We summarize some of the properties of these functions in:

Theorem 6. (The Hellinger transform of a given experiment).

Assume that θ is finite. Let for each prior distribution t on θ , $H(t|E)$ denote the Hellinger transform of E i.e.

$H(t|E) = \int h(dE, t)$ where $h(x, t) = \prod_{\theta} x_\theta^{t_\theta}$ when $x \in [0, \infty]^\theta$. Put also $\tilde{H}(t|E) = \int \tilde{h}(dE, t)$ where $\tilde{h}(x, t) = \prod_{\theta} x_\theta^{t_\theta}$ if $t_\theta > 0$ whenever $x_\theta = 0$ and where $\tilde{h}(x, t) = 0$ otherwise (i.e. if $x_\theta = t_\theta = 0$ for some θ).

Then:

- (i) $H(\cdot|E) > \tilde{H}(\cdot|E)$ for any experiment E .
- (ii) $H(\cdot|E) = \tilde{H}(\cdot|E)$ if and only if E is homogeneous.
- (iii) $H(t|E) = \tilde{H}(t|E)$ for all $t \in K_\theta$ such that $t_\theta > 0$ for all θ .

Furthermore both functions $H(\cdot|E)$ and $\tilde{H}(\cdot|E)$ are log convex. $\tilde{H}(\cdot|\tilde{E})$ is continuous while $H(\cdot|E)$ is continuous if and only if \tilde{E} is homogeneous.

Remark. By theorem 1 $H(\cdot|E) < H(\cdot|F)$ for all $t \in K_\theta$ when $E > F$. In particular $H(\cdot|E) = H(\cdot|F)$ when $E \sim F$ and, by the same theorem, this may be turned around i.e. $H(\cdot|E) = H(\cdot|F)$ if and only if $E \sim F$. By Hansen and Torgersen 1974 it may however easily happen that non comparable linear normal models with unknown variances became comparable by replications. If so and if the experiments are named E and F such that $E \times E > F \times F$ then $H(\cdot|E) = H(\cdot|E \times E)^{\frac{1}{2}} < H(\cdot|F \times F)^{\frac{1}{2}} = H(\cdot|F)$ although $\delta(E, F) > 0$.

References

- Blackwell, D. 1951. Comparison of experiments.
Proc. Second Berkeley Symp. Math. Statist. and Prob.,
pp. 93-102.
- Hansen, O.H. and ^{Torgersen}~~Torgersen~~, E. 1974. Comparison of linear normal
experiments.
Ann. Statist. 2, pp. 367-373.
- Huber, P.L. 1980. Robust statistics. Wiley.
- LeCam, L. 1986. Asymptotic methods in statistical decision theory.
Springer Verlag.
- Morse, N. and Sacksteder, R. 1966. Statistical isomorphism.
Ann. Math. Statist. 37, pp. 203-214.
- Roberts, A.W. and Varberg, D.E. 1973. Convex functions.
Academic Press.
- Torgersen, E. 1976. Comparison of statistical experiments.
Scand. J. Statist. 3, pp. 186-208.
- Torgersen, E. 1981. Measures of information based on comparison
with total information and with total ignorance.
Ann. Statist. 9, pp. 638-657.
- Torgersen, E. 1985. Majorization and approximate majorization for
families of measures.
In Lecture notes in statistics, no. 35, Springer Verlag,
pp. 247-310.